Jing Wang, Dong Yang, Kaimin Chen & Xiaodong Sun (2021) Cruise dynamic pricing based on SARSA algorithm, Maritime Policy & Management, 48:2, 259-282.

Cruise dynamic pricing based on SARSA algorithm

Abstract: It is a common practice to promote highly discounted fares by cruise companies to enlarge the market share, ignoring economically sustainable development. In some regions, the continuous discounted fares leading to the unsatisfying revenue may be the main cause of decline of ports calls. Cruise companies have learned that dynamic pricing would be much more advantageous at revenue management instead of blindly lowering fares. This paper illustrates such an attempt. We try to dynamically price multiple types of staterooms with various occupancies and evaluate the effect on demand and revenue from different discount and refund policies. We first formulate the cruise pricing problem as Markov Decision Process and Reinforcement Learning (RL), more specifically, state-action-reward-state-action (SARSA) algorithm, is applied to solve it. We then use empirical data to validate the feasibility of RL. Results show that both revenue and demand could be improved under reasonable discount policies. In addition, we demonstrate that reasonable refund policies can also facilitate revenue growth. Finally, a comparison between SARSA algorithm and *Q*-learning algorithm is discussed. Our finding suggests that SARSA results in higher revenues but takes more time to converge.

Keywords: Cruise industry; Refund policy; Discount policy; Dynamic pricing; Reinforcement Learning

1. Introduction

In the intense competition of cruise market, promoting highly discounted fares is the common practice for cruise companies to occupy the regional market share. However, the continuous discounted fares leading to the unsatisfying revenue have forced some cruise companies to give up the current market. For example, in Mainland China, the policy of discounted fares to enlarge the market share is one of the accounted reasons for the horning development around the decade (Sun, Ye, and Xu 2016). From 2010 to 2017, about 301.7% increased cruise capacity had been attracted into the Chinese cruise market. However, the long-term lower pricing policy is not economically sustainable for cruise companies, as well as the development of cruise industry. In 2018, a 15.84% decline of ports calls occurred and was followed by a 21% drop in 2019 compared to the previous year accordingly (CLIA 2019). The cruise vessels' deployment out of Chinese market leading to the recession of cruise market has highlighted the need of revenue management (RM) among cruise companies for sustainable development. RM, originated in the airline industry in the 1970s, has been identified an effective method to improve revenue in many industries, such as cargo, hotel, and car rental; but its application is limited in cruise (X. Sun, Jiao, and Tian 2011; Espinet Rius 2018). Dynamic pricing is an active and reliable methodology of RM. In this paper, we illustrate its application to improve cruise revenue, taking into account the characteristics of the cruise industry, which includes unique cabin classes, discount policies, and refund service.

Different from airline and hotel industries, the types of staterooms in a cruise ship are classified in terms of view and occupancy, i.e., interior stateroom, ocean-view stateroom, balcony stateroom, and deluxe suite, each accommodating two, three, or four passengers. Cruise fare is sensitive to the size of traveling parties; an additional person means more potential on-board consumption and higher revenue. As a common practice of cruise companies, a stateroom is priced for at least two passengers and discount is offered to the third and fourth passenger to increase the sales of triple and quadruple staterooms. In this paper, we study three main types of staterooms with three occupancies and five discount policies for each combination of room type and occupancy.

In addition, we address refund service, which was rarely observed in previous literature. Refund service arises from passenger uncertainty prior to departure, due to long booking periods (over a year). It is necessary for alleviating passenger worries and for facilitating demand. Here, we propose and test hierarchical penalty rates under different refund scenarios. Using computational experiments, we are able to examine how refund service affects revenue and demand.

Essentially, our setup is a stochastic optimization problem. Factoring in the combination of room type and occupancy, discount policies, refund service, and the dynamic sales environment, this problem faces what can be deemed as “curse of dimensionality” (Abhuit Gosavi, Bandla, and Das 2002). To tackle this, we use Markov Decision Process (MDP) to formulate cruise dynamic pricing and adopt a developed optimization method, namely, Reinforcement Learning (RL) (McGill and van Ryzin 1999). RL has been increasingly applied in stochastic environments where demand is not well defined (Abhuit Gosavi, Bandla, and Das 2002; Lu, Hong, and Zhang 2018). Cruise demand fluctuates widely with seasons, routes, and products (e.g., activities on board and on shore). In addition, potential passenger reservation price (i.e. willingness to pay) can influence demand which is impossible to be understood in advance (Rana and Oliveira 2014), so as the stochastic demand itself. In previous literature, reservation prices and parameters in demand functions are always assumed such that the problem becomes more tractable. However, these assumptions may cause the computing outcome to deviate from reality (Ding et al. 2020) and cannot reflect the changing environment (e.g., competition and itinerary change), further leading to revenue loss. This paper uses RL to obtain the approximated optimal pricing policies in order to maximize revenue, because it can learn the dynamic relationship of demand and price in the ever-changing sales environment (Gosavi, Bandla, and Das 2002; Lu, Hong, and Zhang 2018). More specifically, State-Action-Reward-State-Action (SARSA), a model-free algorithm of RL, will be employed.

We hope to achieve the following contributions:

(1) We take into consideration the unique operational factors of a cruise ship, including stateroom types with different occupancies, discount policies, and refund service. Our model enriches the literature of cruise dynamic pricing and can provide valuable insights to cruise companies.

(2) This paper is the first and a preliminary attempt to apply RL in cruise dynamic pricing problems. The result reveals that an actual cruise voyage can improve its revenue by up to 26.86% under certain pricing scenarios.

The rest of this paper is organized as follows. Section 2 reviews the literature on dynamic pricing in the cruise industry and the application of RL. The proposed models are introduced in Section 3. SARSA algorithm of RL is detailed in Section 4. Numerical analysis based on historical data is conducted in Section 5. Conclusion and discussion on future work are presented in Section 6.

2. Literature Review

This paper is related to two areas of literature: first, RM of the cruise industry; and second, the application of RL in RM of the other industries.

2.1 Cruise RM

As one of the fastest growing industries, cruise has received increasing research attention from several aspects, e.g., characteristics of cruise trips (Ahmed et al. 2002; Toh, Rivers, and Ling 2005), cruise passenger satisfaction and loyalty (Petrick 2005; Chua et al. 2015), passenger consumption onboard or onshore (Grace Wang et al. 2020), cruise ports competition (Pallis, Arapi, and Papachristou 2019; Jeong and Hyun 2019), etc. However, RM has rarely been discussed. Over the decades, it has been considered as an efficient strategy to improve revenue and has been widely applied in aviation and hotel industries (Klein et al. 2020). The cruise industry has its unique properties, such as cabin classification, extended booking period, etc.

(Neil Biehn 2006; Maddah et al. 2010). Taking them into consideration, we believe it'll be interesting to investigate how RM can benefit the cruise industry.

The existing literature on cruise RM roughly fall into two categories: 1) capacity control and 2) dynamic pricing (pricing optimization). We found four in the former, where the constraints of cabin capacity and lifeboat capacity are examined (Biehn 2006; Ji and Mazzarella 2007; Maddah et al. 2010; Li 2014). Biehn (2006) formulated a deterministic linear program for revenue maximization, limited by the number of cabins and lifeboats. It could be used for accepting or rejecting passenger orders. Maddah et al. (2010) developed a discrete dynamic capacity control model, considering lifeboat capacity and cabin types with the method of dynamic pricing heuristics. Ji and Mazzarella (2007) proposed a nested class allocation of the cruise inventory and a dynamic class allocation adapted from the airline industry. Based on their simulation, revenue could be improved by 4.2%–6.3%. Li (2014) studied the cruise overbooking problem, applying real options. It shows that its proposed overbooking level could decrease the loss of revenue caused by no-shows or cancellations.

The later -- dynamic pricing (pricing optimization) -- is a business strategy that adjusts product prices in a timely fashion in order to allocate the right service to the right customer at the right time (Lin 2006). Our study falls into this category and aims at adjusting cruise fares timely to sell the inventory of different stateroom types over the selling horizon.

Three studies are found in cruise dynamic pricing. Ayvaz-Cavdaroglu, Gauri and Webster (2019) studied customer choice among multiple alternatives differentiated by itineraries, cabin types, and departure dates. Marketing expenses were taken into consideration as well. They applied a multinomial logit choice model, in which assuming the relationship of price and demand was linear. According to the simulation with historical booking data, revenue could be increased by 8%-20%. Focusing on customer arrival rate and remaining sales time and tickets, Dong, Jia, and Xie (2014) formulated a dynamic nonlinear programming problem that assessed the impact of aforementioned three factors on pricing and revenue. The simulation result illustrated that the relationship of price and the three factors was in accordance with the actual situation. Li, Miao, and Wang (2014) modelled cruise RM, considering cabin types with different occupancies. They developed an integer programming model with liner constraints, which can facilitate demand forecasting. With the constraints of non-decreasing fare policy, the proposed model obtained an improved revenue up to 28%. Our study is close to Li, Miao, and Wang (2014) since we also address cabin classification and occupancy. They set discrete fixed-price tiers. Unlike them, we dynamically price cruise fares by the pair of cabin type and occupancy, while treating demand as stochastic. We will also address the manifold discount policies that are widely practiced by cruise companies and evaluate their impacts on revenue.

Overall, this study improves previous literature mainly from two aspects. First, when adopting model-based approaches to solve RM problems, most work demands very specific, and unrealistic, assumptions. For example, Maddah et al. (2010) formulated an MDP framework to address dynamic capacity control, in which the transition probabilities between states were assumed known. Once sales environment changes, pricing policy could deviate from the optimal one to a large extent. Second, when accommodating stochastic demand and multiple operational factors (such as stateroom types, overbooking or refunding service, etc.) in one model, the optimization problem tends to be complicated by large-scale computation. It is known as 'curse of dimensionality', which hinders calculation efficiency (Maddah et al. 2010). Therefore, an approach which is model-free and adapted to large-scale calculations is desired and we choose SARSA algorithm, a traditional RL approach.

2.2 RL application

Gosavii, Bangla, and Das (2002) first proposed RL approach for airline RM; since then, it has received increasing attention in solving large-scale RM problems when demand is non-stationary with unknown characteristics. Although limited in cruise industry, considerate studies have been found in others with perishable products, e.g., aviation, hotel, electricity, etc. Similar to cruise tickets, unsold perishable products will have no salvage value after a deadline. Gosavii, Bandla, and Das (2002) proposed *Q*-learning, another RL approach, to solve the single leg airline problem with multiple fare classes and overbooking. They concluded that RL outperformed heuristic methods, e.g., Nested Expected Marginal Seat Revenue, which was widely used in the airline industry. Rana and Oliveira (2015) solved the optimal pricing problem for perishable interdependent products with RL and analysed the performance of *Q*-learning with eligibility traces algorithm. Lu, Hong, and Zhang (2018) used *Q*-learning to solve the dynamic pricing problem for energy management in a hierarchical market, resulting in profit growth. In these studies, RL and *Q*-learning have shown to be excellent in solving for approximated optimal policies in dynamic pricing problems. Due to the different control mechanisms in the artificial intelligence community (Collins and Thomas 2011), *Q*-learning is denoted as an "off-line" policy, which learns the maximum revenue from existing pricing policies with exploitation ability. Considering the fact that travelling demand is random and occasional, we will employ SARSA, another tradition algorithm of RL, which acts as an "on-line" control method (Sutton and Barto 1998). It has both exploitation and exploration abilities to learn the maximum revenue in a larger explorative range. A comparison between the two algorithms will be presented in Section 5.4.

3. Model Description

The goal of dynamic pricing is to maximize a cruise voyage's revenue through selling a fixed and finite inventory before departure. Cruise companies always set a long selling horizon, generally one year. Fares are dynamically priced in term of remaining inventory (stateroom types plus occupancies) and time before departure. In this study, the dynamic pricing problem acknowledges the stochastic demand and works with the inventory of one cruise voyage. The variables and parameters in this model can be found in Table 1.

Table 1. Variables and parameters in Section 3.

Variables	
i	Number of occupants staying in one stateroom
j	Stateroom type by view
V	Vector of stateroom types in one cruise ship
v^{ij}	Capacity of stateroom type j accommodating i persons (onwards we use stateroom ij for short)
t	Discrete time slot, $t \in [1, T]$
T	Last selling time slot
t'	Purchasing time slot of a later refunded ticket
A_t^{ij}	Price of stateroom ij at time slot t
N_t^{ij}	Potential Demand of stateroom ij at time slot t
p_t^{ij}	Purchasing probability of stateroom ij at time slot t according to passenger reservation price
λ_t^{ij}	Expected potential demand for stateroom ij at time slot t
s_t^{ij}	Remaining inventory for stateroom ij at time slot t
$x_{t't}^{ij}$	Number of refund tickets of staterooms ij, paid at time slot t' but refunded at t
D_{jz}	Discount for the zth passenger accommodated in stateroom type j, $z = 3,4$.
μ_w	w^{th} hierarchical penalty rate of a refund
Decision variables	
a_t^{ij}	Individual ticket price for stateroom ij at time slot t
Parameters	
b^{ij}	Passenger price sensitivity for stateroom ij

3.1. Model assumptions

The process of dynamic pricing can be activated by three events over the selling horizon: 1) a potential passenger arrives, 2) an order is accepted or denied, 3) a refund occurs. To model these events in a Markov Decision Process (MDP), we propose the following assumptions.

(1) The selling horizon is divided into ordered discrete time slots t, each with independent demand. The time slots in this paper are denoted as months both in the actual case and simulated scenarios.

(2) As most passengers take cruise in groups, we only consider staterooms accommodating at least two. V is denoted as the vector of stateroom ij which has j types of

staterooms, each accommodating i passengers. The capacity of each type is represented by v^{ij}. $V = \{v^{ij}\}$, $i = 2,3,4.$, $j = 1,2,3, \ldots, m$.

(3) A stateroom is priced with at least two passengers. It is a common practice to offer the third and fourth passenger extra discounts D_{jz} to attract more people in one stateroom. Stateroom price A_t^{ij} can be expressed in terms of individual price a_t^{ij} as shown in Eq. (1). a_t^{ij} is the decision variables in our model.

$$A_t^{ij} = 2a_t^{ij} + \sum_{z=3}^{4} a_t^{ij} D_{jz}\,, \; z = 3,4. \; j = 2,3,4. \tag{1}$$

(4) As the leisure trips, it is more difficult to handle an overbooked passenger in the cruise voyage in practice (Talluri and Van Ryzin 2005; Ji and Mazzarella 2007). From this view, we don’t consider the overbooking factor in the paper.

(5) We assume the passengers booking different stateroom types are subject to independent distribution. As a result, no substitution is considered, as well as upgrading class.

(6) The fare price which the passengers willing to pay is assumed independent among passengers (Rana and Oliveira 2014). We assume the purchasing probability p_t^{ij} is determined by the fare price. It follows an exponential distribution function as $p_t^{ij} = e^{-b^{ij}A_t^{ij}}$ (Lau and Lau 1988), where b^{ij} is passenger price sensitivity to stateroom type ij (Rana and Oliveira 2015). A_t^{ij} is denoted as the price of stateroom ij at time slot t which is determined by the number of i persons in it.

(7) Potential demand N_t^{ij} follows poison distribution with parameter λ_t^{ij}, which denotes the expected potential demand of stateroom type ij (Gosavi, Bandla and Das, 2002; Rana and Oliveira, 2014).

$$P\{N_t^{ij} = k\} = \frac{\left(\lambda_t^{ij}\right)^k e^{-\lambda_t^{ij}}}{k!}, k = 0,1,2, \ldots, n. \tag{2}$$

(8) Refund service is also considered in this study. We assume that refund tickets follow uniform distribution (Gosavi, Bandla and Das, 2002) , which are generated randomly from the sold tickets. Additionally, we suppose the refund tickets can not exceed the thirty precents of sold tickets to make the simulation more reasonable. The cruise company will charge a hierarchical penalty rate μ_w, related to the time left to departure. In general, the closer to departure refund happens, the higher the penalty. Its hierarchy can be described as $(\mu_1, \mu_2, \dots, \mu_w)$.

(9) Revenue is assumed to come from ticket sales exclusively. The variable cost of one cruise voyage is mainly composed of fuel cost, port fees, employee salaries, and maintenance cost, all of which do not change with the number of passengers on board. Although food and supply costs can fluctuate, it is negligible compared to the total revenue.

3.2. Revenue Model

Based on the assumptions above, we formulate the cruise revenue model as follows:

$$max \sum_{j=1}^{m} \sum_{i}^{2,3,4} \sum_{t=1}^{T} p_t^{ij} N_t^{ij} A_t^{ij} \quad (3)$$

$$s.t. \quad \sum_{j=1}^{m} \sum_{i}^{2,3,4} \sum_{t=1}^{T} p_t^{ij} N_t^{ij} \leq v^{ij} \quad (4)$$

In Eq. (3), A_t^{ij}, the price of stateroom ij, is determined by individual prices a_t^{ij}, the decision variable. $\mathrm{p}_t^{ij}\mathrm{N}_t^{ij}$ is the actual demand which indicates the reservations at time slot *t* based on the purchasing probability p_t^{ij} and the potential demand N_t^{ij}. Eq. (4) states that the actual demand throughout the time slots cannot exceed the capacity of stateroom ij, v^{ij}.

When refund service is considered, the objective function can be derived as Eq. (5).

$$max \sum_{j=1}^{m} \sum_{i}^{2,3,4} \sum_{t=1}^{T} p_t^{ij} N_t^{ij} A_t^{ij} - \sum_{t=1}^{T} \sum_{j=1}^{m} \sum_{i}^{2,3,4} \sum_{t'=1}^{t-1} (1-\mu_w) A_t^{ij} x_{t't}^{ij} \quad (5)$$

$$s.t. \quad \sum_{j=1}^{m} \sum_{i}^{2,3,4} \sum_{t=1}^{T} p_t^{ij} N_t^{ij} \leq v^{ij} \quad (6)$$

$$s_t^{ij} = s_{t-1}^{ij} - p_t^{ij} N_t^{ij} + \sum_{t'=1}^{t-1} x_{t't}^{ij} \quad (7)$$

In Eq. (5), the revenue includes two parts. The first part is the same as Eq. (3), the second part is the loss of revenue due to refund. $x_{t't}^{ij}$ denotes the number of refund tickets of staterooms ij, paid at time slot $t^{'}$ but refunded at t. The refunded tickets need to be added into the current inventory at time slot t. The remaining inventory s_t^{ij} of stateroom ij is updated in Eq. (7). Next, we use RL approach to solve this model.

4. Reinforcement Learning (RL)

We have formulated this cruise dynamic pricing problem as an MDP with discrete time slots over a finite horizon. SARSA, a traditional RL algorithm, will be employed to find the solution, because it is skilled in reaching approximated optimization for large-scale and complicated MDPs (Rana and Oliveira, 2014). To learn the solution, SARSA algorithm runs through trial and error and through interacting with the environment to acquire the approximated optimal pricing policy.

The variables and parameters in this model are listed in Table 2.

Table 2. Variables and parameters of RL in Section 4.

Variables	
s_t^{ij}	Remaining inventory for stateroom ij at time slot t (state in RL)
$\boldsymbol{S}^{ij}$	The set of states representing the remain inventory for stateroom ij
a_t^{ij}	Individual ticket price for stateroom ij at time slot t (action in RL)
$\boldsymbol{A}(s_t^{ij})$	The set of prices can be picked up when there is s_t^{ij} remaining inventory
$\boldsymbol{R}(s_t^{ij})$	Expected reward of stateroom ij from time slot t to T
$r(s_t^{ij})$	Immediate expected reward of stateroom ij at time slot t (reward in RL)
$Q(s_t^{ij}, a_t^{ij})$	Pricing policy when adopting action a_t^{ij} in state s_t^{ij} at time slot t
$Q^*(s_t^{ij}, a_t^{ij})$	Approximated optimal pricing policy with the pairs of a_t^{ij} and s_t^{ij}
$\pi\ (S, a)$	tickets' pricing policy
$\pi^*(S, a)$	Approximated optimal tickets' pricing policy
Parameters	
α	Learning rate
δ	Convergence constant

4.1. Markov Decision Process (MDP)

In MDP framework, the dynamic pricing process can be described as one agent interacting with the environment in order to reach maximum revenue. The agent here is a cruise company that sets fares. The environment is made up with fluctuant demand, discount policies, refund service, etc. The interaction between the agent and the environment is conducted in terms of sequential decision-makings with the following three key components: (1) state s_t^{ij} represents the remaining inventory for stateroom ij at time slot t, $s_t^{ij} \in \boldsymbol{S}^{ij}$, $\boldsymbol{S}^{ij}$ is denoted as the set of states representing the remain inventory for stateroom ij at time t. (2) action a_t^{ij} represents the individual ticket price for stateroom ij at time slot t, $a_t^{ij} \in \boldsymbol{A}(s_t^{ij})$, $\boldsymbol{A}(s_t^{ij})$ is denoted as the set of prices can be picked up when there is s_t^{ij} remaining inventory at time t. (3) reward $r(s_t^{ij})$ represents the immediately expected revenue at time slot t, $\boldsymbol{R}\left(s_t^{ij}\right)$ is the expected revenue when at s_t^{ij}.

For each time slot $t \in \{1,2,\cdots,T\}$. The agent adjusts fares at each time slot t, which is a calendar month in this model. In state s_t^{ij} (remaining inventory), when the agent executes an action a_t^{ij} (individual ticket fare) and then potential passengers in the environment react, the demand $p_t^{ij}N_t^{ij}$ in Section 3.2 is generated. The system will move to a new state s_{t+1}^{ij} (updated ticket inventory) which is $s_t^{ij} - p_t^{ij}N_t^{ij}$. If considering the situation of refund, s_{t+1}^{ij} *should be* $s_{t-1}^{ij} - p_t^{ij}N_t^{ij} + \sum_{t'=1}^{t-1} x_{t't}^{ij}$. The agent achieves a reward $r(s_t^{ij})$ (immediate ticket revenue) and evaluates a_t^{ij}. The reward can be calculated through Eq. (8) or Eq. (9) according to whether refund service is offered. In one episode run, the process will iterate to the last selling time slot T and calculate the cumulative $r(s_t^{ij})$ following the pricing policy.

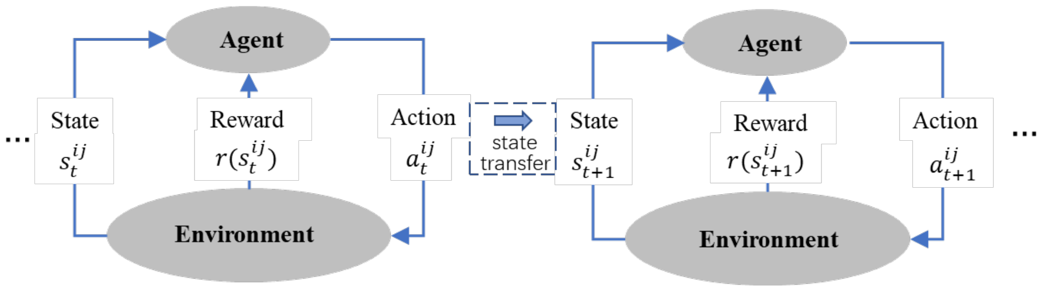

Figure 1. Interaction of agent and environment in MDP model.

At time slot t, the immediate reward $r(s_t^{ij})$ is calculated by the multiplication of the reservations $p_t^{ij} N_t^{ij}$ and the stateroom price A_t^{ij} which is determined by individual price a_t^{ij} in Eq. (1).

$$r(s_t^{ij}) = \sum_{j=1}^{m} \sum_{i}^{2,3,4} p_t^{ij} N_t^{ij} A_t^{ij} \tag{8}$$

If we heed refund service, $r(s_t^{ij})$ is the fare revenue minus the refund.

$$r\left(s_t^{ij}\right) = \sum_{j=1}^{m} \sum_{i}^{2,3,4} p_t^{ij} N_t^{ij} A_t^{ij} - \sum_{j=1}^{m} \sum_{i}^{2,3,4} (1-\mu_w) A_t^{ij} x_{t't}^{ij} \tag{9}$$

The objective of the MDP model is to maximize the expected total revenue, starting from s_t^{ij} at time slot t, i.e., $R\left(s_t^{ij}\right)$, also the cumulative sum of $r(s_t^{ij})$ from t to T as in Eq. (10).

$$\boldsymbol{R}\left(s_t^{ij}\right) = r\left(s_t^{ij}\right) + r\left(s_{t+1}^{ij}\right) + \cdots + r\left(s_T^{ij}\right) \tag{10}$$

To obtain the maximum $\boldsymbol{R}\left(s_t^{ij}\right)$, SARSA algorithm will be applied and, in the meantime, generate the approximated optimal pricing policy $\pi^*(S,a)$. We will illustrate how it works in Section 4.2.

4.2. **SARSA** *algorithm*

SARSA algorithm, a temporal difference method of RL, can solve MDP models by evaluating state-action values through iterations. In SARSA, $Q\left(s_t^{ij}, a_t^{ij}\right)$ denotes the state-action value function that represents the system's future total reward from setting a price a_t^{ij} at inventory level s_t^{ij} and thereafter generates a pricing policy $\pi\,(S,a)$. Q-value obtained by $Q\left(s_t^{ij}, a_t^{ij}\right)$

is assigned to evaluate the level of expected revenue $R(s_t^{ij})$. SARSA algorithm follows an updating mechanism for a larger Q-value until the maximum Q-value appears. In the learning process, the previous $Q(s_t^{ij}, a_t^{ij})$ is stored and updated by the current $Q(s_{t+1}^{ij}, a_{t+1}^{ij})$ through iterations until $Q^*(s_t^{ij}, a_t^{ij})$ is obtained. The approximated optimal pricing policy $\pi^*(S, a)$ can be extrapolated from $Q^*(s_t^{ij}, a_t^{ij})$. $\pi^*(S, a)$ takes the form of $\pi^*\left((s_1^{ij}, a_1^{ij}), (s_2^{ij}, a_2^{ij}), \dots , (s_T^{ij}, a_T^{ij})\right)$, which indicates the optimal set of fares $\boldsymbol{A}(s_t^{ij})$ that brings the maximum revenue. The mechanism of SARSA is illustrated in Table 3.

Table 3. ***SARSA*** **algorithm.**

Step 1: Initialization
Generate $Q(s_t^{ij}, a_t^{ij})$ arbitrarily, **for** t =1, 2, 3, …, T
Step 2: Iteration
At each time slot t in state s_t^{ij}, select and execute an action a_t^{ij}
Obtain immediate reward $r(s_t^{ij})$
while $t > 0$ **do**
    Step 2.1: Iteration in episode $\boldsymbol{l}$
    if $t \neq T$ **then**
        Move to the next state s_{t+1}^{ij}
        Exploit an action a_{t+1}^{ij} by ε-greedy policy, $Q(s_{t+1}^{ij}, a_{t+1}^{ij})$ obtained
        Update $Q(s_t^{ij}, a_t^{ij})$
        $Q(s_t^{ij}, a_t^{ij}) \leftarrow Q(s_t^{ij}, a_t^{ij}) + \alpha[r(s_t^{ij}) + \gamma Q(s_{t+1}^{ij}, a_{t+1}^{ij}) - Q(s_t^{ij}, a_t^{ij})]$
        repeat until T
    else $t = T$ **then**
        if $\left|Q^l\left(S_T^{ij}, a_T^{ij}\right) - Q^{l-1}\left(S_T^{ij}, a_T^{ij}\right)\right| < \delta$ **then**
            Obtain the optimal policy $\pi^*(S, a) = arg\,max Q^*(s_T^{ij}, a_T^{ij})$
            break
        else if $\left|Q^l\left(S_T^{ij}, a_T^{ij}\right) - Q^{l-1}\left(S_T^{ij}, a_T^{ij}\right)\right| \geq \delta$ **then**
            Step 2.2: iteration in next episode $\boldsymbol{l+1}$
            Move to the next episode $l + 1$
            repeat the Step 2.1
        End
    End
end

In Step 1, $Q(s_t^{ij}, a_t^{ij})$ can be initialized as 0 (Sutton and Barto 1998) or experience from the real-world system as simulator (Gosavi 2009), for example, historical ordering data from a cruise voyage. Then the system moves to iterations in Step 2 for the optimal pricing policy $\pi^*(S, a)$. Two elements are essential in the iteration progress: the ε-greedy policy and the learning rate α. The ε-greedy policy is introduced to guarantee global optimum instead of a local optimum, indicating the system will scan almost all feasible fares for maximum revenue. Given ticket inventory s_t^{ij}, the agent explores a new and unknown fare with the probability $1-\varepsilon \in [0,1]$ (Lu, Hong, and Zhang 2018) or exploits from the existing fare list with the probability ε. A new and larger Q-value determined by $\left(s_t^{ij}, a_t^{ij}\right)$ pair may be generated and stored for the following iteration. In term of updating rule, a random $Q\left(s_{t+1}^{ij}, a_{t+1}^{ij}\right)$ will be chosen with ε-greediness as well to update $Q\left(s_t^{ij}, a_t^{ij}\right)$ as shown in Eq. (11). The ε-greedy policy reinforces SARSA algorithm by updating in a wider range with more randomness. The learning rate α indicates the rate of convergence but irrelevant to the result. The bigger α is, the newer trials override the old ones, which indicates more time the agent takes to interact with the environment, but less time spent to explore new actions to increase Q-value.

$$Q\left(s_t^{ij}, a_t^{ij}\right) \leftarrow Q\left(s_t^{ij}, a_t^{ij}\right) + \alpha\left[r\left(s_t^{ij}\right) + \gamma Q\left(s_{t+1}^{ij}, a_{t+1}^{ij}\right) - Q\left(s_t^{ij}, a_t^{ij}\right)\right] \tag{11}$$

In episode l, when update reaches final time slot T, the system will verify whether it has converged to the maximum Q-value. The maximum Q-value is compared to the previous Q-value obtained in episode $l-1$ to verify if their difference converges to a very small constant δ. δ is the termination criterion as shown in Eq. (12).

$$\left|Q^l\left(s_T^{ij}, a_T^{ij}\right) - Q^{l-1}\left(s_T^{ij}, a_T^{ij}\right)\right| \leq \delta \tag{12}$$

If not, the system will move to episode $l+1$ for iterations. Once converged, the maximum Q-value will be obtained, the approximated optimal pricing policy $\pi^*(S,a)$ can be extrapolated in the terminal iteration through Eq. (13). We set δ as 0.01. Its convergence has been well demonstrated in literature (Jaakkola, Jordan, and Singh 1994) . Figure 2 shows the entire iteration process for one stateroom type.

$$\forall t \in (1,T),\ \forall s_t^{ij} \in \boldsymbol{S}^{ij},\ \pi^*(S,a) = arg\,max Q^*\left(s_T^{ij}, a_T^{ij}\right),\ a_t^{ij} \in \boldsymbol{A}\left(s_t^{ij}\right) \tag{13}$$

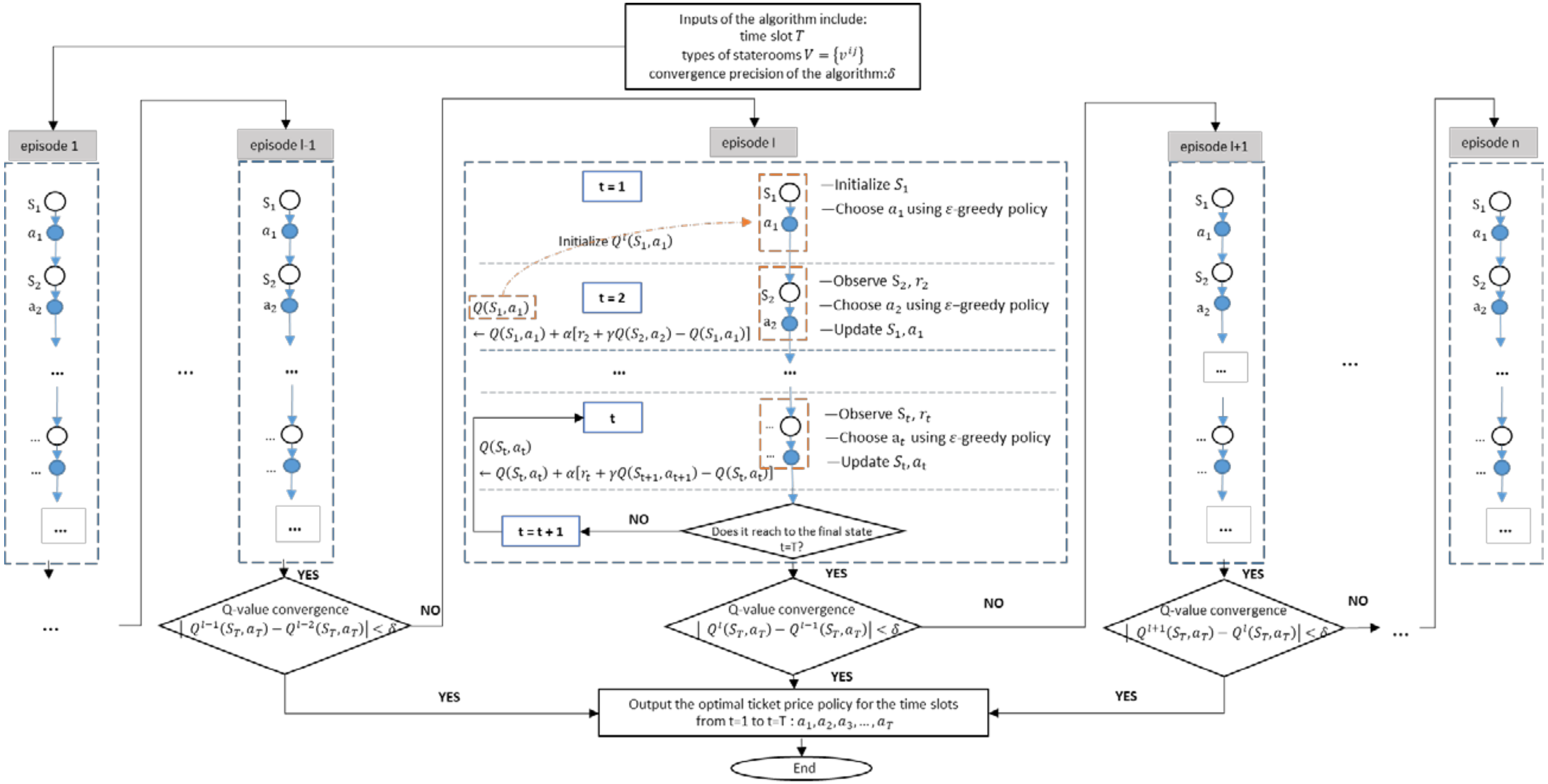

Figure 2. Flowchart for implementing SASAR algorithm.

5. Scenario Analysis

In this section, we report the results from a set of computational experiments. They are based on one actual cruise voyage and we use the result to evaluate the effect of discount policies and refund service on cruise RM. The actual case and the simulated scenarios are described in Section 5.1. The performance of SARSA algorithm is presented in Section 5.2. We also illustrate the differences between SARSA and *Q*-learning algorithms in their solving performance and convergent rates. Section 5.3 discusses the impact of discount policies and refund service on cruise RM, divided by stateroom type and occupancy.

5.1. Scenarios description

5.1.1. Actual case

The actual case is a real cruise voyage operated by a North American cruise company. The information is investigated from its private database which records the reserved data of the previous voyages including voyage profiles, passengers' information, reserved time and prices. We are allowed to utilize the data for RM analysis. We choose one voyage randomly, running a typical seven-day itinerary in the western Caribbean from the collected database which contains over 80 thousand pieces of reservations of one cruise ship over 50 voyages. The selling horizon was about one year, which was split into 12 time slots equally in our study. The cruise ship can accommodate 2,063 passengers in 867 staterooms. There are six stateroom types: balcony stateroom, interior stateroom, ocean view stateroom, suite stateroom, portholes stateroom, and upper/lower stateroom. We find the prices of the last three are relatively stable and they only account for 8.72% of the total revenue and 7.07% of the total capacity. Therefore, we choose to focus on the other three types, i.e., balcony stateroom, interior stateroom, and ocean view stateroom. Each type is further subdivided into double, triple, and quadruple occupancies. Booking data can be found in Table 4.

Table 4. The original information of the actual scenario.

Type of stateroom	Occupant number	Reserved staterooms(room)	Capacity(room)	Load factor
Balcony	double	194	266	72.93%
	triple	28	48	58.33%
	quadruple	27	52	51.92%
Interior	double	188	276	68.12%
	triple	23	40	57.50%
	quadruple	38	48	79.17%
Ocean view	double	81	96	84.38%
	triple	14	17	82.35%
	quadruple	18	24	75.00%
Total		611	867	70.47%

Use double occupancy balcony stateroom as an example, as in Figure 3. Monthly price fluctuated wildly, from 50% discount to nearly full price, with no particular pattern. This indicates that tickets were probably priced randomly without heed of demand and inventory, which can hardly constitute an optimal policy. In the following sections, we will demonstrate the application of RL in cruise RM and, moreover, incorporate discount policies and refund service.

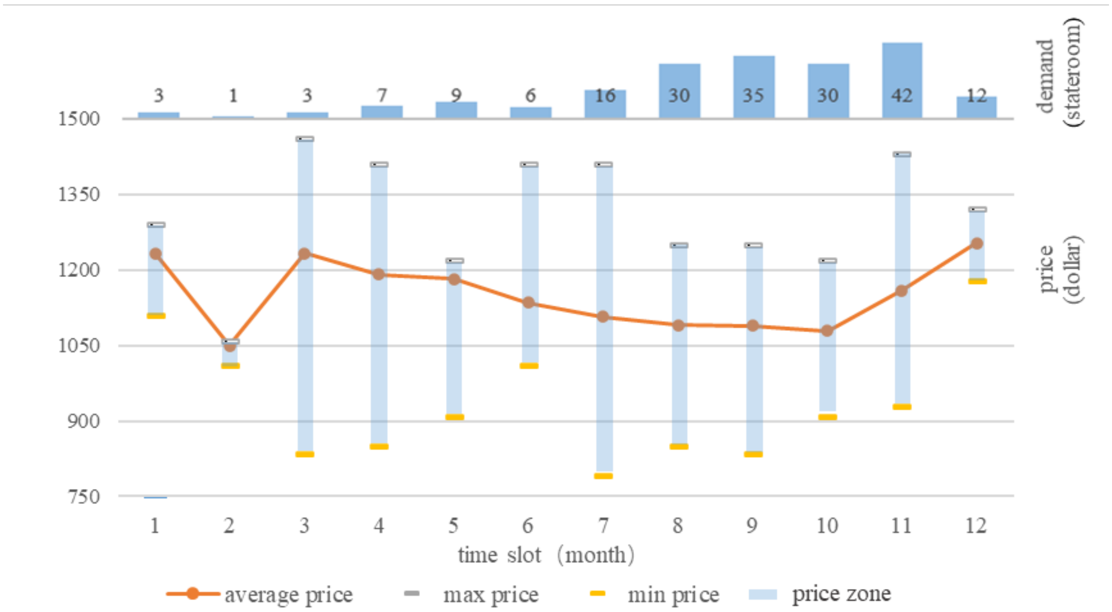

Figure 3. Price and demand of double occupancy balcony in the actual scenario.

5.1.2. Scenarios Simulation

First, we classify five scenarios by a tiered discount policy. It is a common practice to offer discounts to the third and fourth passenger in one stateroom. They are Discount 1 (1, 1, 1), Discount 2 (0.8, 0.8, 0.8), Discount 3 (0.8, 0.7, 0.5), Discount 4 (0.5, 0.5, 0.5), and Discount 5 (0.5, 0.5, 0)). Discount 1 to Discount 5 represent the discount tiers from no discount rising to the deepest discount. Within a discount tier, the first number in the bracket indicates the discount offered for the third passenger in a triple occupancy stateroom, and the second and third numbers are the discounts for the third passenger and fourth passenger, respectively, in a quadruple occupancy stateroom.

Second, we classify four refunding scenarios with different penalty rate μ_w. We depict μ_w as (μ_1, μ_2, μ_3), where μ_1, μ_2, μ_3 denote three penalty rates determined by the remaining time before departure. Generally, more than six months, three to six months, and less than three months are the most used time windows in practice. The four refunding scenarios can be expressed as (0%, 0%, 0%), (50%, 50%, 50%), (50%, 80%, 100%), (80%, 80%, 80%). Take (50%, 80%, 100%) for example, a customer will lose 50% of the fare if she cancels her reservation six months before departure; she will be charged 80% of the fare when seeking refund between three and six months prior and get no money back if less than three months before departure.

In the simulation, for each type of stateroom, we set a minimal ticket price found in the collected data as the lower bound. In the dynamic pricing process at each time slot, fares can fluctuate between the lower bound and full prices in the increment of 10 dollars. Based on Arevalillo (2019), we have accessed the price sensitivities b^{ij} from the previous over 80 thousand pieces of reservations in the database, which is shown in Table 5. The purchasing probability p_t^{ij} is calculated by the accessed $b^{ij}$and pricing information A_t^{ij} in the actual cruise voyage. In the simulation, the expected potential demand λ_t^{ij} in the simulation is estimated by the actual demand divided by p_t^{ij}. The potential demand N_t^{ij} in the simulation can be obtained by the poison distribution in Eq. (2), which can be found in 'potential demand' column in Table 6. The actual demand can also be found in Table 6. We can observe the difference of actual demands between the actual scenario and simulated scenarios.

Table 5. Price sensitivity b^{ij}

Occupant number / Stateroom type	Double	Triple	Quadruple
Balcony	1/3500	1/5000	1/5500
Interior	1/2500	1/2800	1/3800
Ocean view	1/1800	1/2400	1/350

5.2. Algorithm Performance

5.2.1. SARSA Performance

We use Python 3.7.1 to simulate all the scenarios on a Windows PC with Intel Dual Core processors, i7-10710U CPU, 1.6GHz and16G RAM.

The approximated optimal policies are derived by averaging 100 independent simulation runs ($\alpha = 0.05$, $\gamma = 0.8$, $\delta = 0.01$), among which 93% can converge in an average of 25000 iterations. The average running time is about 15'30 minutes.

Figure 4 shows the convergence of balcony, interior and ocean view staterooms separately in one run. It illustrates that the model can quickly learn to get larger *Q*-values through trial and error in about 5000 iterations and converge to the maximum *Q*-value in about 2×10^4 iterations, which means the approximated optimal pricing policy is obtained.

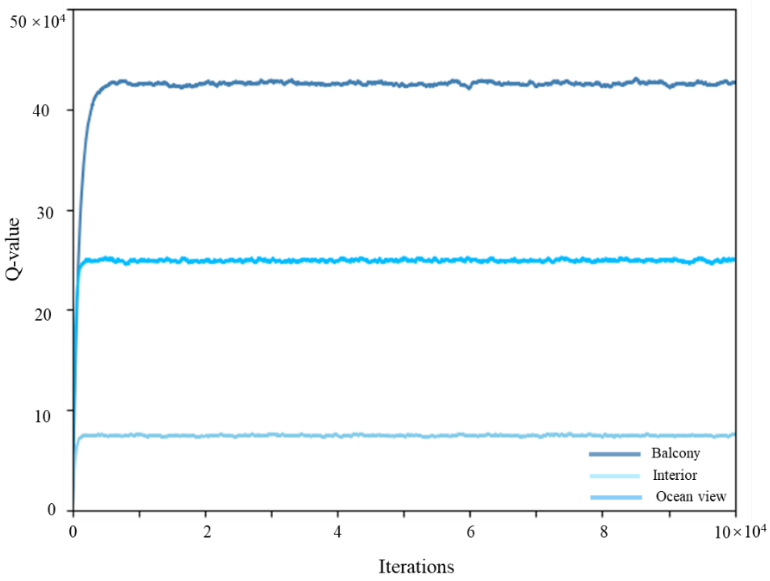

Figure 4. Convergence of *Q*-values for three types of staterooms.

5.2.2. Q-learning vs. SARSA

Q-learning, another RL algorithm, has been broadly applied to address MDP problems; similar to SARSA, it also implements a ε-greedy policy. But the two algorithms follow different procedures to update current *Q*-values (A Collins and L Thomas 2011). ε-greedy policy is a method of choosing an action to update to a larger *Q*-value until the maximum revenue is obtained. More specifically, it either explores a new action with the possibility of ε or exploits an existing optimal action with the possibility of $1-\varepsilon$. Unlike SARSA that updates *Q*-value by exploring new actions, *Q*-learning algorithm updates *Q*-value through exploiting the best policy so far in order to maximize the expected revenue as shown in Eq. (14). Obviously, *Q*-learning may miss the opportunity of a better and unknown policy, but it makes the updating process more efficient.

$$Q(s_t^{ij}, a_t^{ij}) \leftarrow Q(s_t^{ij}, a_t^{ij}) + \alpha[r(s_t^{ij}) + \gamma maxQ(s_{t+1}^{ij}, a_{t+1}^{ij}) - Q(s_t^{ij}, a_t^{ij})] \quad (14)$$

Both algorithms have converged in the simulation. *Q*-learning is faster and converges in 17,352 iterations, in contrast to 24,235 iterations of SARSA. The enlarged areas in Figure 5 show that the convergence curve of SARSA fluctuates more than that of *Q*-learning. This is due to the fact that SARSA is always exploring new and unknown actions, rather than exploiting the existing maximum Q-value as in Q-learning. In another word, SARSA "loves" learning more than Q-learning. SARSA outperforms Q-learning with a slightly higher revenue of $753,196 versus $737,250 from Q-learning. Overall, they both are proved efficient in solving our dynamic pricing problems but boast different advantages: SARSA can achieve higher revenue but converge slower than Q-learning.

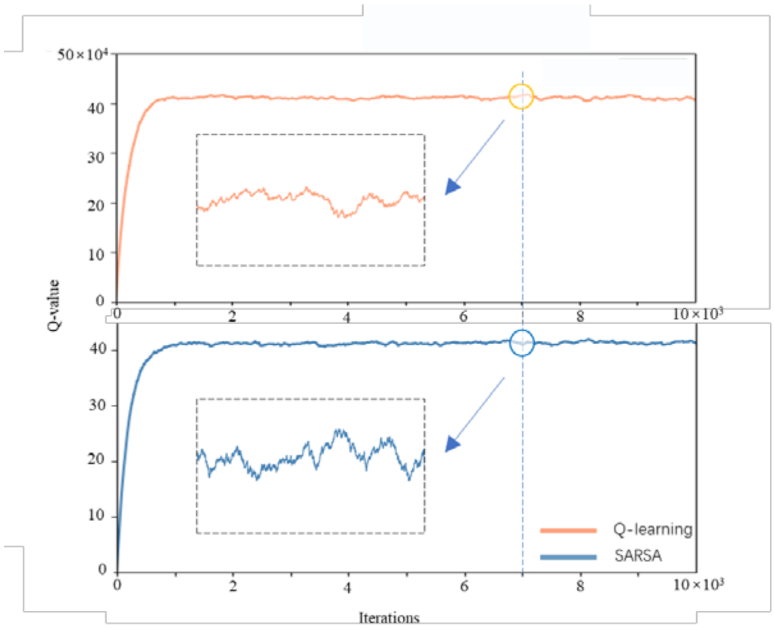

Figure 5. Comparison between *Q*-learning and SARSA algorithms.

5.3. Result Analysis

5.3.1. Discount policies analysis

In section 5.1.2, we set five discount scenarios by discount policies. After simulation, we obtained the approximated optimal pricing policies for each proposed scenario. Figure 6 shows the simulated approximately optimal prices of different staterooms with different discount throughout the time slots compared to the original prices.

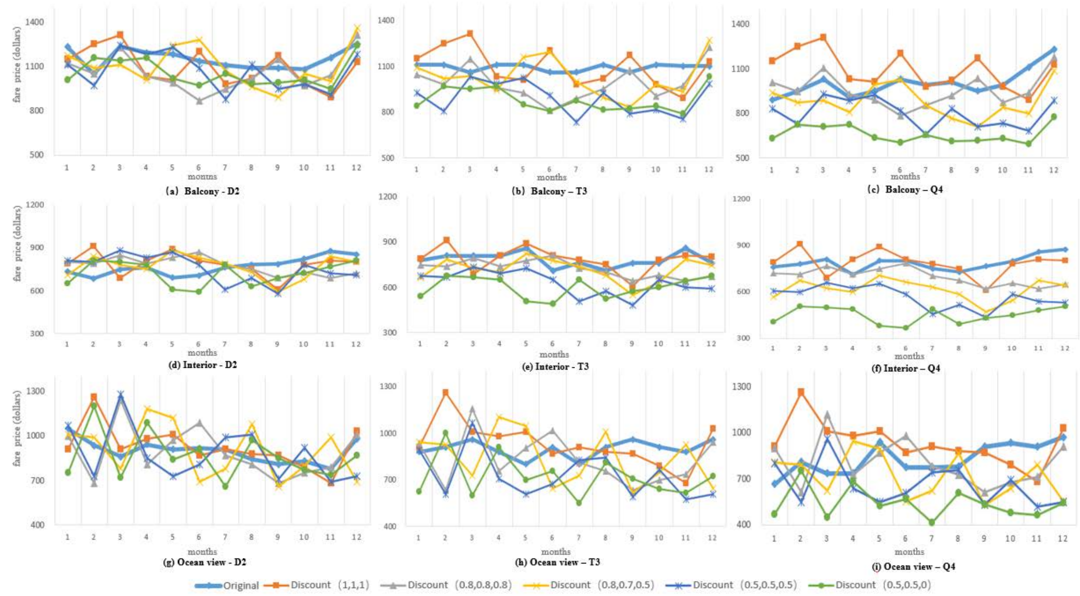

Figure 6. Simulated approximately optimal prices of different staterooms with different discount throughout the time slots.

Overall, the simulated optimal prices appear more volatile than the relatively stable original prices. Compared with the three different types of staterooms (balcony, interior and ocean view) accommodating the same number of passengers, the prices of ocean view stateroom type show a widest fluctuation. This may be explained by the smallest capacity and its highest price sensitivity, leading to the price reacting more strongly to the change of state. In terms of occupants in the stateroom, the discount policy has more significant influence on the quadruple occupancy stateroom, followed by the triple and double ones.

In Table 6, in the five discount scenarios, every number in a coloured block indicates the change in demand from our simulation result to its corresponding value in the actual case. Besides Discount 1 (1, 1, 1), which indicates no discount, most of the demands increase. Also notice that sales peak arrives earlier, around May to July, than in the actual scenario, September to November. An early sales peak allows cruise companies more flexibility to adjust their sales plans.

First, among the three stateroom types, ocean view stateroom is always the most popular choice for passengers and has a relatively higher load factor than the other two types. This may be explained by its limited supply (137 staterooms) as to the other two (730 staterooms). A discount will aggravate passengers' anxiety of this stateroom type being sold out, leading to more sales. Secondly, balcony staterooms enjoy boosted sales most from the percentage of increase in demand; it increases sharply from Discount 3 (0.8, 0.7, 0.5) and reaches to 92.59% with Discount 4 (0.5, 0.5, 0.5), where the cruise ship has achieved a full capacity.

In terms of occupancy, first, though double staterooms come with no discount, it is interesting to find that discount policies have a positive effect on their sales, i.e., increase of 7.73%-22.68% for balcony staterooms, 19.15%-30.85% for interior staterooms, and 2.27%-13.58% for ocean view staterooms. Second, triple occupancy sales always see a small increase. This can probably be attributed to its rigid target market of three-member families that cannot be divided into double occupancy staterooms or integrated into quadrable occupancy staterooms. Finally, quadruple occupancy staterooms can be influenced greatly by the discount policies, i.e., the load factor can be 100% under Discount 4 (0.5, 0.5, 0.5) and Discount 5 (0.5, 0.5, 0). The reason might be the highest passenger price sensitivities among the three occupancies.

Table 6. Comparison of load factors and demand rising rates between the original and the discount scenarios.

Stateroom type	Month	Potential demand			Actual demand			Discount 1 (1,1,1)			Discount 2 (0.8, 0.8, 0.8)			Discount 3 (0.8, 0.7 ,0.5)			Discount 4 (0.5,0.5,0.5)			Discount 5 (0.5,0.5,0)		
		D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4
Balcony	1	5	4	2	3	2	1	2	-1	2	2	1	0	2	2	1	0	4	1	1	2	1
	2	2	2	2	1	1	1	0	-1	0	2	0	-1	3	0	-1	4	0	1	3	1	3
	3	6	4	2	3	2	1	10	-1	0	2	-1	-1	0	-1	3	1	-1	1	4	-1	1
	4	12	2	6	7	1	3	4	0	-3	-1	0	3	8	0	-2	4	0	-2	10	0	-1
	5	16	2	8	9	1	4	4	3	-2	3	0	1	0	2	4	10	1	1	10	2	-1
	6	11	6	6	6	3	3	10	4	-1	17	-2	3	9	0	2	14	-3	2	27	-1	4
	7	29	15	2	16	8	1	5	-2	0	8	-3	0	2	-6	-1	11	-1	9	3	-5	3
	8	56	8	10	30	4	5	-9	-2	-1	0	2	3	-2	-3	9	-13	1	-3	2	2	6
	9	62	4	6	35	2	3	-7	-1	-1	-6	3	-1	-3	5	3	-1	0	2	-1	5	8
	10	55	8	2	30	4	1	5	-3	2	5	4	0	4	2	0	-5	2	2	-14	1	5
	11	82	0	7	42	0	3	-5	3	-2	1	3	5	5	4	2	6	1	5	0	6	-3
	12	25	0	2	12	0	1	-4	2	-1	0	3	0	4	4	0	11	16	6	-1	8	-1
	Total	361	54	56	194	28	27	15	1	-7	33	10	12	32	9	20	42	20	25	44	20	25
Load factor		-	-	-	73.21%	58.33%	51.92%	78.57%	60.42%	38.46%	83.83%	60.42%	53.85%	85.34%	79.17%	75.00%	88.72%	100.00%	100.00%	89.47%	100.00%	100.00%
Increasing rate of		-	-	-	-	-	-	7.73%	3.57%	-25.93%	17.01%	35.71%	44.44%	16.49%	32.14%	74.07%	21.65%	71.43%	92.59%	22.68%	71.43%	92.59%
Stateroom type	Month	Potential demand			Actual demand			(1,1,1)			(0.8, 0.8, 0.8)			(0.8, 0.7 ,0.5)			(0.5,0.5,0.5)			(0.5,0.5,0)		
		D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4
Interior	1	4	2	2	2	1	1	1	0	1	0	1	1	1	1	0	0	2	0	1	1	1
	2	3	2	2	2	1	1	1	0	2	3	1	-1	5	0	-1	1	0	1	5	-1	0
	3	5	2	2	3	1	1	1	1	0	1	2	-1	3	-1	0	-1	1	3	-1	1	0
	4	11	2	2	6	1	1	3	0	0	3	0	-1	3	2	0	1	0	0	2	0	1
	5	7	5	0	4	2	0	1	1	0	4	-1	0	1	2	0	0	2	0	12	0	2
	6	11	4	5	6	2	2	6	-1	3	1	3	-2	4	0	-1	9	-1	0	11	0	1
	7	35	2	11	19	1	5	10	-1	-2	4	0	2	5	-1	1	9	0	1	15	2	8
	8	52	11	11	28	5	5	4	-4	-2	6	0	-1	-2	3	0	7	-1	4	5	-1	2
	9	71	7	9	38	3	4	7	0	-1	-1	-1	0	9	1	8	-3	4	2	-5	4	7
	10	60	5	14	31	2	6	7	-1	-3	2	2	2	5	0	2	8	3	9	5	2	0
	11	50	5	12	25	2	5	8	-1	-3	11	-1	3	6	-1	1	14	2	-3	6	2	-5
	12	48	5	18	24	2	7	-13	-1	-5	1	0	1	6	0	-2	3	0	-7	-10	5	-7
	Total	357	53	88	188	23	38	36	-7	-10	35	6	3	46	6	8	48	12	10	46	15	10
Load factor		-	-	-	68.12%	57.50%	79.17%	81.16%	40.00%	58.33%	81.16%	40.00%	81.25%	80.80%	72.50%	85.42%	84.78%	87.50%	100.00%	85.51%	95.00%	100.00%
Increasing rate of		-	-	-	-	-	-	19.15%	-30.43%	-26.32%	18.62%	26.09%	7.89%	24.47%	26.09%	21.05%	25.53%	52.17%	26.32%	24.47%	65.22%	26.32%
Stateroom type	Month	Potential demand			Actual demand			(1,1,1)			(0.8, 0.8, 0.8)			(0.8, 0.7 ,0.5)			(0.5,0.5,0.5)			(0.5,0.5,0)		
		D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4	D2	T3	Q4
Ocean view	1	3	3	2	1	1	1	1	2	0	2	3	0	1	0	0	-1	3	1	1	3	-1
	2	6	3	2	2	1	1	-1	-1	0	-2	-1	1	0	0	0	1	0	-1	-2	-1	0
	3	3	0	2	1	0	1	0	1	0	3	0	0	1	1	0	1	0	-1	2	1	-1
	4	9	3	2	3	1	1	-2	0	0	1	0	-1	0	1	0	-1	-1	1	3	0	0
	5	5	5	2	2	2	1	-1	-2	1	4	1	-1	-1	-1	1	0	0	0	-1	0	1
	6	22	3	4	8	1	2	-3	0	-1	-7	0	0	-3	1	0	-4	0	-1	-5	0	0
	7	5	3	4	2	1	2	4	0	-1	5	-1	-1	7	1	-1	6	-1	1	11	-1	4
	8	28	3	6	11	1	3	-1	0	-1	0	-1	2	-1	-1	-1	-3	1	0	-6	-1	4
	9	22	7	2	9	2	1	1	-1	0	6	-2	1	8	-1	-1	5	1	2	7	0	0
	10	38	6	4	15	2	2	3	-2	0	7	-1	1	2	-1	1	0	-1	2	-3	-1	2
	11	48	3	4	20	1	2	1	0	-1	7	1	0	-7	0	0	3	2	0	0	1	-2
	12	21	3	2	7	1	1	0	0	0	-3	1	-1	3	0	4	4	-1	2	2	2	-1
	Total	210	43	36	81	14	18	2	-3	-3	8	0	1	10	0	3	11	3	6	9	3	6
Load factor		-	-	-	84.38%	82.35%	75.00%	86.46%	64.71%	62.50%	86.46%	82.35%	66.67%	92.71%	82.35%	79.17%	95.83%	100.00%	100.00%	93.75%	100.00%	100.00%
Increasing rate of		-	-	-	-	-	-	2.47%	-21.43%	-16.67%	9.88%	0.00%	5.56%	12.35%	0.00%	16.67%	13.58%	21.43%	33.33%	11.11%	21.43%	33.33%

-20

50

D2-double occupancy stateroom;T3-triple occupancy stateroom; Q4-quadruple occupancy stateroom

Figure 7 shows the obtained revenues in the five discount scenarios. It shows that balcony staterooms contribute more to revenue while the input from ocean view staterooms appears minimal. The revenue from balcony and ocean view staterooms displays a climbing trend from Discount 2 (0.8, 0.8, 0.8) to Discount 4 (0.5, 0.5, 0.5). For interior staterooms, the revenue declines slightly in Discount 4 but is still higher than the original scenario. All three stateroom types show a slight decrease in revenue in Discount 5 (0.5, 0.5, 0).

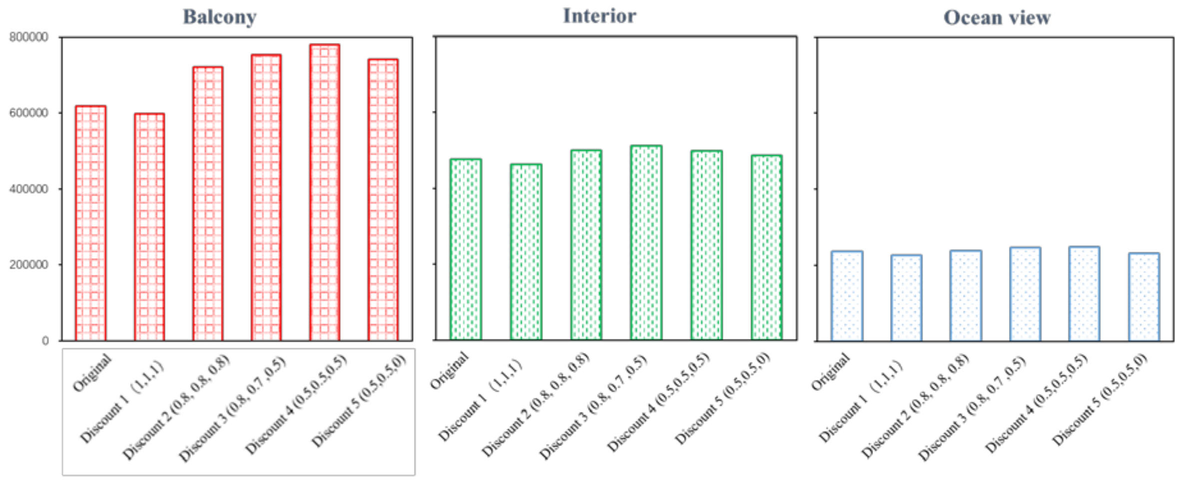

Figure 7. Revenue from each stateroom type under different discount scenarios.

Figure 8 displays the percentage of increase in revenue (from the actual case) under Discount 1 (1, 1, 1) to Discount 5 (0.5, 0.5, 0). Balcony stateroom has a sharper increase than the other two types. The revenue growth under Discount 4 (0.5, 0.5, 0.5) once jumps to 26.26%. Interior stateroom can obtain the highest increase with Discount 3 (0.8, 0.7, 0.5). Though Discount 5 (0.5, 0.5, 0), the deepest discount policy, leads to the highest load factors in all staterooms, it brings revenue loss.

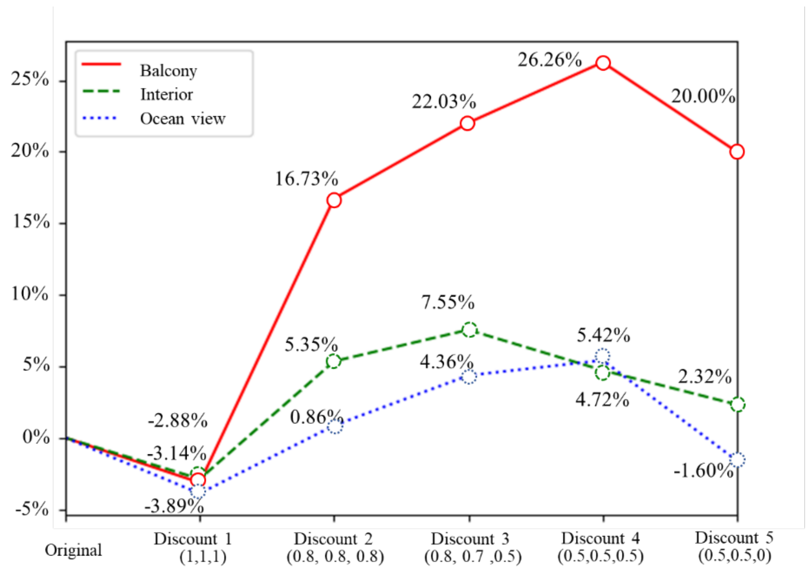

Figure 8. Increasing rates of revenue for each type of stateroom under diverse discount scenarios.

5.3.2. Refund analysis

We evaluate whether refund service is beneficial to cruise RM, specifically with four sets of penalty rates ((0%, 0%, 0%), (50%, 50%, 50%), (50%, 80%, 100%), (80%, 80%, 80%)).

First, we examine refund service's influence on different stateroom types. In the simulation, the refunded tickets for different staterooms accommodating double, triple and quadruple passengers can be observed in Figure 9.

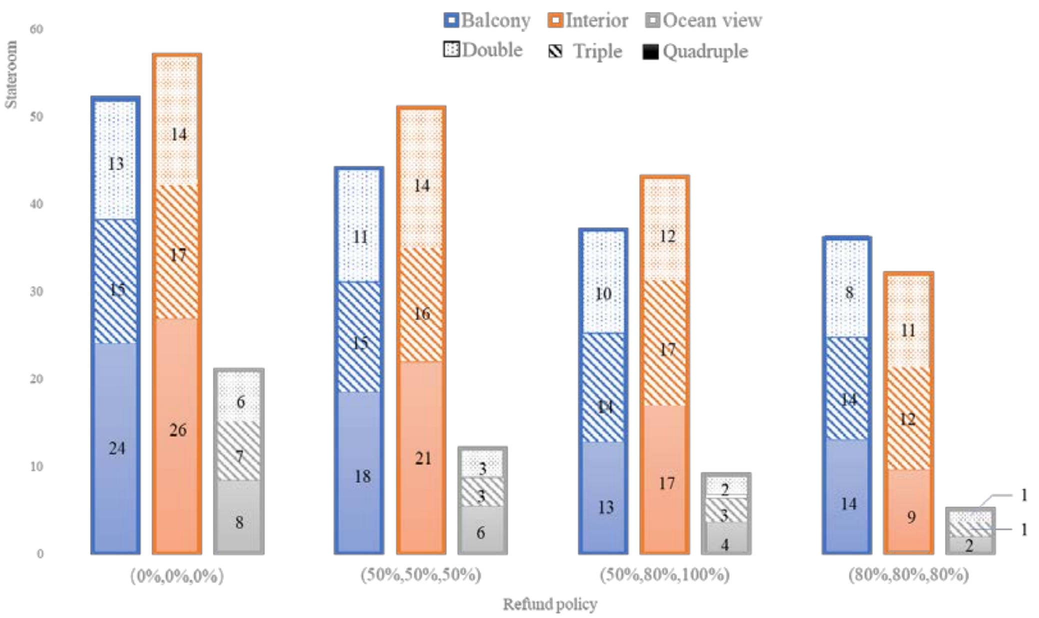

Figure 9. The refunded tickets $x_{t't}^{ij}$ of different statements *ij* under different discounts

In Section 3.1 (7), we assume the refund tickets follow uniform distribution. The number of refunded tickets $x_{t't}^{ij}$ is randomly created in the simulation. The pricing procedure with the refunded service is to update new state, generating different prices which influence the revenue. From this view, we have focused on the change of demand and revenue under different refunded policy in order to reflect the effect of refund service which is shown in Figure 10.

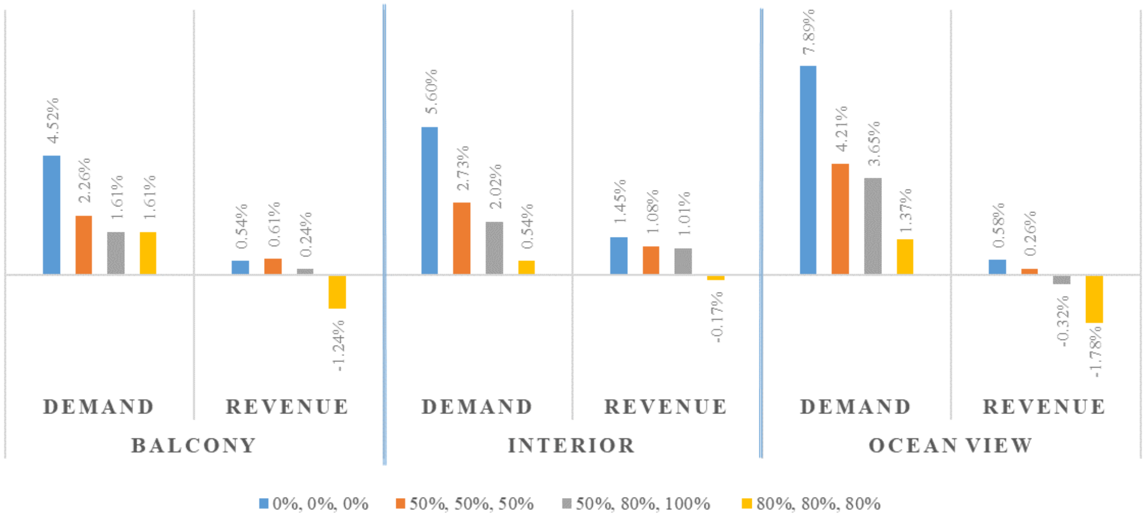

Figure 10. The effect of refund service on different stateroom types

The total revenue is almost always increased (from the actual case), from 0.26% to 1.45%, except for a couple of cases under (50%, 80%, 100%) and (80%, 80%, 80%). This indicates the forgiving refund policies can improve the total revenue due to the partial penalty obtained by the refunded behaviours and the revenue of re-selling capacity relinquished by refunded passengers. Additionally, it has a positive effect on the demand as a result of the adjustment of fare price at a new remaining inventory with re-selling tickets.

Then, we will examine the effects of refund on occupancies. Take balcony stateroom as an example and compound refund with the three discount policies for analysis, Discount 2 (0.8, 0.8, 0.8), Discount 3 (0.8, 0.7, 0.5) and Discount 4 (0.5, 0.5, 0.5), representing high discounts, hierarchical discounts, and low discounts respectively.

For double occupancy staterooms, refund policies (penalty rates) of (0%, 0%, 0%), (50%, 50%, 50%), (50%, 80%, 100%) have a positive effect on both demand and revenue, shown in Figure 9. The refund policy of (80%, 80%, 80%) can slightly improve the demand but lead to revenue loss. The demand and the revenue of triple and quadruple occupancy staterooms fluctuate more than double ones. This indicates that the refund service is more beneficial for higher-occupancy staterooms. Under Discount 4 (0.5, 0.5, 0.5), the refund service has obvious negative impact on the revenue and demand of triple and quadruple occupancy staterooms. In the scenario of low discounts, passengers are more likely to accept the preferential discounted fares than refunds.

When considering the total demand and revenue of balcony staterooms (double, triple and quadruple occupant staterooms together) under different discount scenarios, we can observe that the refund service performs better under Discount 2 (0.8, 0.8, 0.8), followed by Discount 3 (0.8, 0.7, 0.5). The demand growth rates increase from -3.45% to 10.71% under Discount 2 (0.8, 0.8, 0.8) and from 0 to 8.51% under Discount 3 (0.8, 0.7, 0.5). The revenue is improved from -2.06% to 10.30% under Discount 2 (0.8, 0.8, 0.8) and -5.5% to 8.60% under Discount 3 (0.8, 0.7, 0.5). Under Discount 4 (0.5, 0.5, 0.5), only (0%, 0%, 0%) and (50%, 50%, 50%) refund policies can grow the demand and revenue slightly.

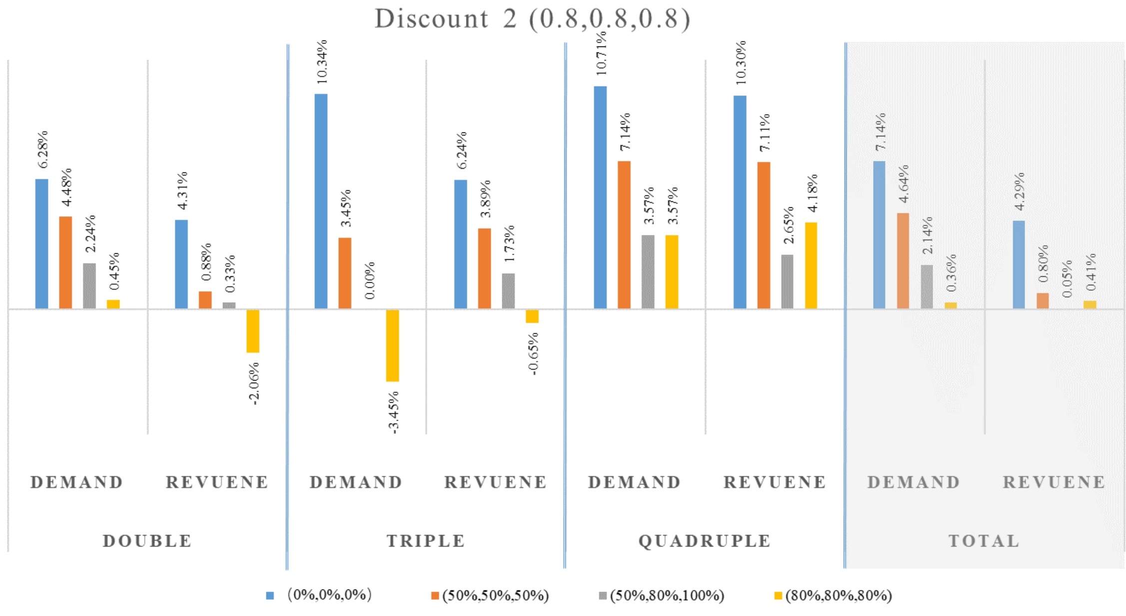

Figure 9 (a). Effect of refund on different occupancy staterooms under Discount 2 (0.8, 0.8, 0.8).

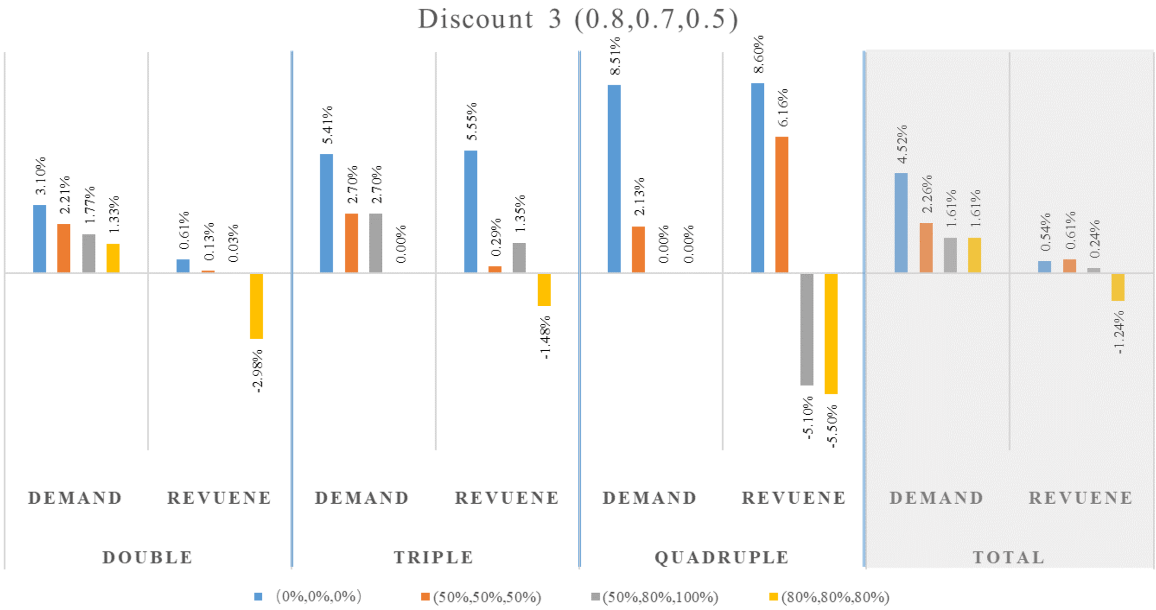

Figure 9 (b). Effect of refund on different occupancy staterooms in Discount 3 (0.8, 0.7, 0.5).

Figure 9 (c). Effect of refund on different occupancy staterooms in Discount 4 (0.5, 0.5, 0.5).

According to the results of our computational experiments, certain refund policies (e.g. (0%, 0%, 0%), (50%, 50%, 50%), (50%, 80%, 100%)) can increase revenue, attracting more demand. The results also suggest that refund service is more favourable in lower discount scenarios. Compared to double occupancy staterooms, triple and quadruple ones receive a greater impact from the refund service. This analysis can provide cruise managers with some sales suggestions.

6. Conclusion and Discussion

The recession of cruise market has heightened the necessity for cruise RM. This paper aims to provide a dynamic pricing model to maximize revenue for cruise companies. We formulate this problem as a finite MDP, taking into consideration the stochastic demand and the characteristics of cruise services (i.e., types of staterooms, occupancy, discount policies, and refund service). SARSA algorithm of RL approach is applied to solve this large-scale complex optimization problem. The relationship of states (inventories) and actions (prices) in the MDP can be learned through ongoing simulation instead of being assumed as in the precious literature. RL proves to be feasible for implementing RM in the cruise industry that faces the uncertainty of demand and flexible fares. Simulation results show that it can improve the revenue up to 26.26% with discount policies and 4.29% with refund service.

This study can provide cruise managers with useful pricing insights. In different discount scenarios, stateroom sales of different types and occupancies react distinctively: such as, balcony staterooms are the main contributor to revenue growth; the discount policies of triple and quadruple occupancy staterooms have a positive effect on the sales of double occupancy ones. In this study, we also examine refund service which was rarely addressed before. It is interesting to find that, in certain refund scenarios, although there exists partial loss in fares, both demand and the total revenue can be improved. In particular, under lower discount scenarios, the refund policy implemented on double occupancy staterooms has a stronger effect on the revenue compared to it adopted on triple and quadruple occupant staterooms, while this effect is minor under high discount scenarios.

In the future, we would like to continue studying cruise RM and confront the limitations of this research: (1) Though we considered multiple types of staterooms, upgrade among different types has not been addressed, which may further improve revenue. (2) Cruise companies tend to use aggressive discount policies as a means to encourage on-board consumption. Therefore, it is necessary to investigate the trade-off between the two in RM. We believe both are well worth pursuing in the future.

References

A Collins, and L Thomas. 2011. "Comparing Reinforcement Learning Approaches for Solving Game Theoretic Models: A Dynamic Airline Pricing Game Example." *Journal of the Operational Research Society* 63 (8): 1165–73. https://doi.org/10.1057/jors.2011.94.

Arevalillo, Jorge M. 2019. "A Machine Learning Approach to Assess Price Sensitivity with Application to Automobile Loan Segmentation." *Applied Soft Computing* 76 (March): 390–99. https://doi.org/10.1016/j.asoc.2018.12.012.

Ayvaz-Cavdaroglu, Nur, Dinesh K. Gauri, and Scott Webster. 2017. "Empirical Evidence of Revenue Management in the Cruise Line Industry." *Journal of Travel Research* 58 (1): 104–20. https://doi.org/10.1177/0047287517737178.

Ding, Yi, Kaimin Chen, Dongming Xu, and Zhang Qiong. 2020. "Dynamic Pricing Research for Container Terminal Handling Charge." *Maritime Policy & Management*, July, 19. https://doi.org/10.1080/03088839.2020.1790051.

Dong, Ya Ru, Bin Jia, and Dong Fan Xie. 2014. "Data Processing and Modeling of Dynamic Pricing of Cruise Transportation Based on Revenue Management." *Advanced Materials Research* 1030–1032 (September): 1984–87. https://doi.org/10.4028/www.scientific.net/AMR.1030-1032.1984.

Espinet Rius, Josep Mª. 2018. "Global and Local Pricing Strategies in the Cruise Industry." *Revenue & Pricing Management*, June, 12. https://doi.org/10.1057/s41272-018-00155-5.

Gosavi, Abhijit. 2009. "Reinforcement Learning: A Tutorial Survey and Recent Advances." *INFORMS Journal on Computing* 21 (2): 178–92. https://doi.org/10.1287/ijoc.1080.0305.

Gosavi, Abhuit, Naveen Bandla, and Tapas K. Das. 2002. "A Reinforcement Learning Approach to a Single Leg Airline Revenue Management Problem with Multiple Fare Classes and Overbooking." *IIE Transactions* 34 (9): 729–42. https://doi.org/10.1080/07408170208928908.

Grace Wang, Wen-Huei Chang, Yue Cui, Guanqiu Qi, and Kevin X Li. 2020. "Introducing an Economic Impact Platform to Navigate Cruise Value-Added Chain with Environmental Considerations." *Marine Policy* 112 (February): 103713. https://doi.org/10.1016/j.marpol.2019.103713.

Jaakkola, Tommi, Michael I. Jordan, and Satinder P. Singh. 1994. "On the Convergence of Stochastic Iterative Dynamic Programming Algorithms." *Neural Computation* 6 (6): 1185–1201. https://doi.org/10.1162/neco.1994.6.6.1185.

Jeong, Ji Youn, and Sunghyup Sean Hyun. 2019. “Roles of Passengers’ Engagement Memory and Two-Way Communication in the Premium Price and Information Cost Perceptions of a Luxury Cruise.” *Tourism Management Perspectives* 32 (October): 100559. https://doi.org/10.1016/j.tmp.2019.100559.

Ji, Lu, and Joseph Mazzarella. 2007. “Application of Modified Nested and Dynamic Class Allocation Models for Cruise Line Revenue Management.” *Journal of Revenue and Pricing Management* 6 (1): 19–32. https://doi.org/10.1057/palgrave.rpm.5160065.

Klein, Robert, Sebastian Koch, Claudius Steinhardt, and Arne K. Strauss. 2020. “A Review of Revenue Management: Recent Generalizations and Advances in Industry Applications.” *European Journal of Operational Research* 284 (2): 397–412. https://doi.org/10.1016/j.ejor.2019.06.034.

Lau, Amy Hing-Ling, and Hon-Shiang Lau. 1988. “The Newsbov Problem With Price-Dependent Demand Distribution.” *IIE Transactions*, 9.

Li, Bingzhou. 2014. “A Cruise Line Dynamic Overbooking Model with Multiple Cabin Types from the View of Real Options.” *Cornell Hospitality Quarterly* 55 (2): 197–209. https://doi.org/10.1177/1938965513507126.

Li, Yihua, Qing Miao, and Wang. 2014. “Modeling a Cruise Line Revenue Management Problem.” *Journal of Revenue and Pricing Management* 13 (3): 247–60. https://doi.org/10.1057/rpm.2013.44.

Lin, Kyle Y. 2006. “Dynamic Pricing with Real-Time Demand Learning.” *European Journal of Operational Research* 174 (1): 522–38. https://doi.org/10.1016/j.ejor.2005.01.041.

Lu, Renzhi, Seung Ho Hong, and Xiongfeng Zhang. 2018. “A Dynamic Pricing Demand Response Algorithm for Smart Grid: Reinforcement Learning Approach.” *Applied Energy* 220 (June): 220–30. https://doi.org/10.1016/j.apenergy.2018.03.072.

Maddah, Bacel, Lama Moussawi-Haidar, Muhammad El-Taha, and Hussein Rida. 2010. “Dynamic Cruise Ship Revenue Management.” *European Journal of Operational Research* 207 (1): 445–55. https://doi.org/10.1016/j.ejor.2010.03.043.

Neil Biehn. 2006. “A Cruise Ship Is Not a Floating Hotel.” *Journal of Revenue and Pricing Management* 5 (2): 135–42. https://doi.org/10.1057/palgrave.rpm.5160034.

Pallis, Athanasios A., Kleopatra P. Arapi, and Aimilia A. Papachristou. 2019. “Models of Cruise Ports Governance.” *Maritime Policy & Management* 46 (5): 630–51. https://doi.org/10.1080/03088839.2019.1590657.

Rana, Rupal, and Fernando S. Oliveira. 2014a. “Real-Time Dynamic Pricing in a Non-Stationary Environment Using Model-Free Reinforcement Learning.” *Omega* 47 (September): 116–26. https://doi.org/10.1016/j.omega.2013.10.004.

Rana, Rupal, and Fernando S. Oliveira. 2015. “Dynamic Pricing Policies for Interdependent Perishable Products or Services Using Reinforcement Learning.” *Expert Systems with Applications* 42 (1): 426–36. doi:10.1016/j.eswa.2014.07.007.

Sun, Xiaodong, Yue Jiao, and Peng Tian. 2011. “Marketing Research and Revenue Optimization for the Cruise Industry: A Concise Review.” *International Journal of Hospitality Management* 30 (3): 746–55. https://doi.org/10.1016/j.ijhm.2010.11.007.

Sutton, Richard S., and Andrew G. Barto. 1998. *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning. Cambridge, Mass: MIT Press.